DOCUMENT RESUME

ED 389 738 TM 024 367

AUTHOR Faggen, Jane; And Others
TITLE Effects of Mode of Item Presentation on Standard Setting.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-95-26
PUB DATE Aug 95
NOTE 24p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Assisted Testing; *Cutting Scores; *Difficulty Level; Evaluators; *Pass Fail Grading; Standards; Test Format; *Test Items
IDENTIFIERS Paper and Pencil Tests; *Standard Setting

ABSTRACT

The objective of this study was to determine the degree to which recommendations for passing scores, calculated on the basis of a traditional standard-setting methodology, might be affected by the mode (paper versus computer-screen prints) in which test items were presented to standard setting panelists. Results were based on the judgments of 31 panelists for items in each of 3 parts of a computer-based academic skills assessment designed for use as part of the teacher licensing process. None of the evidence collected--either the analysis of panelists' estimates of item difficulty or their perceptions of effects--revealed any differences in recommended passing standards resulting from the mode of item presentation. (Contains two figures and eight tables). (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

RR-95-26

ED 389 738

RESEARCH REPORT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. I. Braun

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC).

EFFECTS OF MODE OF ITEM PRESENTATION ON STANDARD SETTING

Jane Faggen
Gerald Melican
Donald Powers

BEST COPY AVAILABLE

Educational Testing Service
Princeton, New Jersey
August 1995

TM024367

Effects of Mode of Item Presentation on Standard Setting

Jane Faggen, Gerald Melican, and Don Powers
Educational Testing Service

Copyright © 1995. Educational Testing Service. All rights reserved.

Effects of Mode of Item Presentation on Standard Setting

Jane Faggen, Gerald Melican, and Don Powers
Educational Testing Service

ABSTRACT

The objective of this study was to determine the degree to which recommendations for passing scores, calculated on the basis of a traditional standard-setting methodology, might be affected by the mode (paper vs. computer screen prints) in which test items were presented to standard setting panelists. Results were based on the judgments of 31 panelists for items in each of three parts of a computer-based academic skills assessment designed for use as part of the teacher licensing process. None of the evidence collected--either the analysis of panelists' estimates of item difficulty or their perceptions of effects--revealed any differences in recommended passing standards resulting from the mode of item presentation.

Effects of Mode of Item Presentation on Standard Setting

Jane Faggen, Gerald Melican, and Donald Powers
Educational Testing Service

INTRODUCTION

Computers enable a number of measurement opportunities that are not readily implemented with paper-and-pencil methods, such as computer adaptive testing. However, computerization may also present certain challenges. For instance, when setting passing scores for computer-based tests (CBT), it would seem desirable to present test items to standard setting panelists in the same manner that they are shown to test-takers, i.e., in a computerized format. This ideal may not always be possible, however, because of time pressures and other constraints.

The purpose of the study described here was to determine any differences in recommended passing scores that may result from the mode (paper vs. computer) in which test items are presented to standard-setting panelists. The focus of the study was three computer adaptive tests suitable for assessing basic academic skills in mathematics, reading, and writing of college students who have completed their sophomore year of study. These measures were being considered as part of the teacher certification process in one state.

METHODOLOGY AND MATERIALS

Panel Members

Thirty-one practicing teachers, representative of the teaching pool in the state in terms of gender, ethnicity, teaching-level, area of certification, educational background, and years of experience were selected for the study (see Tables 1A and B). An effort was made to include equal numbers of teachers who had experience with computers and those who had little or no experience. Panel members met on two different days, 15 on the first day and 16 on the second day.

Forms

With the exception of the essay section of the Writing test, the

three computer-based academic skills assessments are computer-adaptive tests. For administration purposes, a fixed number of items is presented to each examinee. The Reading adaptive test presents 36 questions in a 90-minute time-period. The questions may require highlighting information, choosing one or more answers, moving information, or checking boxes in a table. The Writing adaptive test consists of 35 multiple-choice error recognition questions delivered in a 40-minute time-frame. The Mathematics adaptive test, 29 questions in 55 minutes, may require highlighting an answer, moving numbers or symbols, marking a scale, completing a graph, checking boxes in a table, or entering a numerical response.

Each examinee receives a different set of test questions, depending on performance on the previously presented item(s). The scores reported to examinees based on their CBT adaptive test performance are transformed using the 3-parameter IRT model to estimated number right scores on the relevant 40-item reference form, and then linearly scaled to the reporting scale, ranging from 300 to 335. The reference forms represent the mix of content as described by the test specifications. For the purposes of the standard setting study, panel members made judgments on the 40-item reference form for each test. A questionnaire was administered to collect background information from each panel member and information regarding the panelist's computer experience. For the purposes of this study, the three 40-item reference (mathematics, reading, and writing) tests were presented to the standard setting panel members either at the computer terminal or in a screen-print version (i.e., each page is a printed copy of one computer screen).

Data Collection Design

The modified Tucker/Angoff method was used to solicit judgments and establish recommendations for passing scores. Panelists were trained and then asked to estimate the percentage of minimally competent beginning teachers who would know the answer to each question (often referred to as knowledge estimation values) on the 40-item reference form for each test. Half of the items were presented in computer format and half in print-screen (paper) format, according to the design shown in Figure 1. For each item, approximately half of the judgments were made under the computer-

presentation condition and half under the paper-presentation condition.

On the first day, four subpanels were constituted and each reviewed the same test items, but in a different sequence and presentation mode (see Figure 2a). For example, Group I judged mathematics items 1-20 on computer and mathematics items 21-40 in paper format. A second subpanel, Group III, judged the mathematics items in the same order as the first subpanel, but with items 1-20 given in paper format and items 21-40 shown on the computer. The order of the two sets of items was reversed for Groups II and IV. Note that there was a confounding of mode of presentation with order of presentation for the writing items: the paper (screenprint) versions of the writing items were always presented before the computer presented versions. Thus, any effects of order and of mode cannot be disentangled for these items.

On the second day, four new subpanels were constituted, and each reviewed the same groups of 40 test items (see Figure 2b). The design for the first day was followed exactly for the reading and the mathematics items. In order to provide additional time for panelists' transitions from computer stations to worktables, the investigators decided to relax the design requirements for the writing items. Judgments were made based only on the paper versions of the writing items. Furthermore, a perusal of the Day 1 judgments confirmed the *a priori* assumption that the writing items were the least susceptible of any items to a presentation mode effect.

After completing the knowledge estimation judgments, panel members reviewed their paper-based estimates in light of how these same items appeared on the computer and were asked if they would raise or lower any of their previous estimates.

Judgmental Process

After panelists completed recording their knowledge estimation values, they were asked, for each test section, to indicate whether their judgments would have been the same regardless of the mode of presentation. Those who said they would not have been the same were asked to indicate the particular items for which estimates would have differed and to give the reasons for the expected differences.

Analyses

For the purpose of estimating a recommended passing score, the knowledge estimation values are summed for each set of items to calculate an unadjusted recommended passing score. Next, an adjustment is applied to compensate for the possibility that test-takers are likely to answer some questions on the test correctly, not by knowing the answer to the question, but by guessing the correct answer. Summing the adjustments for guessing across all items produces the chance score for these questions. (Note: In most cases, the probability of guessing the correct answer is equivalent to the IRT c-parameter for the item.) The chance score is then added to the unadjusted recommended passing score to produce the recommended passing score. These procedures have been fully described in numerous standard setting papers and reports (see Faggen, 1983, Faggen, et al. 1994, and Livingston & Zieky, 1982). Passing scores were computed twice for each set of reference test items, once on the basis of the paper-based judgments and the other on the basis of the computer-based judgments.

For the purpose of this study, an analysis of variance was carried out for the mathematics and reading tests with day (of data collection), order (item set shown first or second), and mode of presentation (paper vs. computer) identified as the factors. These analyses were based on the unadjusted "passing" scores for each of the sets of reference test items.

RESULTS and DISCUSSION

Table 2 shows the recommended cut scores (unadjusted raw-score study values) for each of two sets of 20 items within each reference test section. Differences due to mode of presentation are small and statistically nonsignificant for each section. For writing items, only the data collected on day 1 were analyzed for the reason noted above. It should be reiterated, however, that the results pertaining to the writing items are confounded in the manner discussed previously (i.e., print-screen items sets were *always* presented prior to the computer versions of the items).

Tables 3 and 4 display, for each set of mathematics and reading

reference test items, the results of the analysis of variance, in which day (of data collection), order (item set shown first or second), and mode of presentation (paper vs. computer) were the factors. All two- and three-way interaction terms were also included in each analysis. Tables 5A and B and Tables 6A and B present three-way tables, Day by Mode by Order, for the mathematics and the reading items; Tables 7A and B and Tables 8A and B contain the respective two-way tables, Day by Mode.

None of the factors or interactions among them accounted for a significant ($p < .05$) portion of variation among cut scores computed from panelists' estimates, although the Day factor for the mathematics Item Set 1 approached significance. The Day 1 panel members consistently generated higher cut scores than the Day 2 panel: Mathematics Items 1-20 (Table 7A) and Items 21-40 (Table 7B); and Reading Items 1-20 (Table 8A) and Items 21-40 (Table 8B). From a practical perspective, these differences are small (in terms of standard deviation units), and of more importance, the recommended cut scores are higher for the Day 1 panel, whether the items were presented via computer or in print-screen (paper) mode. In summary, cut scores were not significantly different according to the mode in which panelists viewed the items. For writing items, a t-test between modes for each set of items revealed no significant difference for either set, $t=0.97$ and $t=1.35$, both of which were less than the selected critical values for $p < .05$.

With respect to panelists' assessments of the effects of mode of presentation, the following was found. The 31 panelists were, according to their questionnaire responses, unanimous in their agreement that the familiarization training they received at the computer was adequate for the standard-setting tasks they were asked to accomplish. Nearly all (28 of 31) of the panelists reported that they were comfortable making judgments about the items presented by computer, regardless of their prior experience with computers. (Only one panelist had never worked with a computer prior to this study, and only six reported that they were not comfortable with a computer.)

Only five of 29 panelists said that their knowledge estimation judgments would have differed according to whether the items were presented in a paper or computer mode. Four of these panelists cited reading items

and one cited both reading and writing items. Only two panelists indicated whether their estimates would have been higher or lower if they had seen the items on the computer instead of on paper. One responded "higher" and the other "lower." Thus, there was little indication that panelists perceived any differences due to the mode of presentation.

CONCLUSION

Neither the results of the analyses of the estimates provided by standard setting panelists nor the opinions of panelists about the possibility of an effect due to mode of item presentation suggest that items need to be presented by computer (rather than as screen prints) in order to obtain credible cut score estimates. The recommendation made here is that screenprints used in the three basic skills assessments of mathematics, reading, and writing can serve as appropriate stimuli for conducting standard setting studies.

This result is a promising one for those who carry out standard setting studies for computer-based tests, since computer equipment is usually not readily available to accommodate the numbers of panel members that are invited to participate in a standard setting study and there is a high cost associated with transporting loaded terminals and screens. However, before this result can be generalized to other knowledge domains and other item-types, suitable replications of this work would need to be carried out.

References

Faggen, J. (1983). *Report on a study of the NTE Core Battery tests by the state of New York*. Princeton, NJ: Educational Testing Service.

Faggen, J., Melican, G., Swartz, R., & Shack, S. (1993, revised January 1994). *Standard setting for Praxis II subject assessments in Hawaii: 16 modules in six subject areas*. Unpublished statistical report. Princeton, NJ: Educational Testing Service.

Livingston, S. & Zieky, M. (1982). *Passing scores*. Princeton, NJ: Educational Testing Service.

FIGURE 1

Basic Study Design

Subpanel	Item Sequence[a]	
	Set 1 (1-20)	Set 2 (21-40)
1 (N≈10)	Computer	Paper
2 (N≈10)	Paper	Computer
	Set 2 (21-40)	Set 1 (1-20)
3 (N≈10)	Computer	Paper
4 (N≈10)	Paper	Computer

Notes Computer = computer presentation of items
Paper = paper presentation of items (print screens)

[a]The item block designated as Set 1 is presented to a subpanel before Set 2.

FIGURE 2a

Data Collection Scheme Day 1

Group I	**Begin at terminal.**			
	Session 1	Computer	Math	1-20	
	Session 2	Paper	Math	21-40	2 back-to-back sessions
	Session 3	Paper	Writing	1-20	with paper (P) copy
	Session 4	Computer	Writing	21-40	
	Session 5	Paper	Reading	21-40	
	Session 6	Computer	Reading	1-20	
Group II	**Begin at terminal.**			
	Session 1	Computer	Math	21-40	
	Session 2	Paper	Math	1-20	2 back-to-back sessions
	Session 3	Paper	Writing	1-20	with paper (P) copy
	Session 4	Computer	Writing	21-40	
	Session 5	Paper	Reading	1-20	
	Session 6	Computer	Reading	21-40	
Group III	**Begin with paper copy of test items.**			
	Session 1	Paper	Writing	21-40	2 back-to-back sessions
	Session 2	Paper	Math	1-20	with paper (P) copy
	Session 3	Computer	Math	21-40	3 back-to back sessions
	Session 4	Computer	Writing	1-20	at terminal (C)
	Session 5	Computer	Reading	1-20	
	Session 6	Paper	Reading	21-40	
Group IV	**Begin with paper copy of test items.**			
	Session 1	Paper	Writing	21-40	2 back-to-back sessions
	Session 2	Paper	Math	21-40	with paper (P) copy
	Session 3	Computer	Math	1-20	3 back-to-back sessions
	Session 4	Computer	Writing	1-20	at terminal (C)
	Session 5	Computer	Reading	21-40	
	Session 6	Paper	Reading	1-20	

FIGURE 2b

Data Collection Scheme Day 2

Group I	Begin at terminal.			
	Session 1	Computer	Math	1-20	
	Session 2	Paper	Math	21-40	4 back-to-back sessions
	Session 3	Paper	Writing	1-20	with paper (P) copy
	Session 4	Paper	Writing	21-40	
	Session 5	Paper	Reading	21-40	
	Session 6	Computer	Reading	1-20	
Group II	Begin at terminal.			
	Session 1	Computer	Math	21-40	
	Session 2	Paper	Math	1-20	4 back-to-back sessions
	Session 3	Paper	Writing	1-20	with paper (P) copy
	Session 4	Paper	Writing	21-40	
	Session 5	Paper	Reading	1-20	
	Session 6	Computer	Reading	21-40	
Group III	Begin with paper copy of test items.			
	Session 1	Paper	Writing	21-40	2 back-to-back sessions
	Session 2	Paper	Math	1-20	with paper (P) copy
	Session 3	Computer	Math	21-40	
	Session 4	Paper	Writing	1-20	
	Session 5	Computer	Reading	1-20	
	Session 6	Paper	Reading	21-40	
Group IV	Begin with paper copy of test items.			
	Session 1	Paper	Writing	21-40	2 back-to-back sessions
	Session 2	Paper	Math	21-40	with paper (P) copy
	Session 3	Computer	Math	1-20	
	Session 4	Paper	Writing	1-20	
	Session 5	Computer	Reading	21-40	
	Session 6	Paper	Reading	1-20	

TABLE 1A

GENDER, ETHNICITY, AND YEARS OF TEACHING EXPERIENCE OF PANEL MEMBERS (N=31)

Gender	Number	Percent
Male	13	41.9%
Female	18	58.1%
Ethnicity		
Black or African American	4	12.9%
Puerto Rican	1	3.2%
White (non-Hispanic)	26	83.9%
Years of Teaching Experience[a]		
1-5	8	25.8%
6-10	2	6.5%
11-15	4	12.9%
16-20	4	12.9%
20+	12	38.7%

[a]Percentages do not sum to 100 because one panelist did not provide this information.

TABLE 1B

TEACHER CERTIFICATION/LICENSES HELD AND COURSES CURRENTLY TAUGHT BY COMPUTER BASED ACADEMIC SKILLS ASSESSMENTS PANEL MEMBERS (N=31)

Teacher Certification/Licenses Held	Currently Teaching
1. Art Education	Art
2. Biology 7-12	Master Teacher - Science Department
3. Business Education 7-12	Business
4. Chemistry/General Science/Math/Spanish/ Professional Administration	Assistant Coordinator Bilingual Department (Chemistry/Math)
5. Comprehensive Special Education, Administration	Supervisor of Learning Disabilities
6. State K-12	Associate Professor of Education
7. State Initial	Gifted and Talented (Pre-K to 6)
8. Education	Third Grade
9. Elementary Education	Elementary School
10. English	English
11. English, History, Social Studies 7-12 Reading Consultant K-12	English 9-12
12. English/Reading/History/Administration	English Department Chair (Grades 7 & 8)
13. English 7-12, Intermediate Administration and Supervisor	Director of Academic Resource Center, Assistant Professor of English
14. French/Italian 7-12	District Department Chair - Foreign Languages
15. Grades 4-8 and 9-12/Social Studies	Elementary School
16. History	History
17. History/Social Studies 7-12	Social Studies
18. Home Economics K-12	Home Economics 7 & 8
19. Math/Physics/Chemistry/General Science	Physics/General Science
20. Mathematics	Mathematics
21. Mathematics/Professional Certificate	Math 11-12 (Pre-Calculus, 11th Honors, Trig)
22. Music Education	Music K-12
23. No Certificate Listed	Chair Mathematical Sciences at xxx State University
24. Physical Education	Physical Education - Elementary
25. Physics/General Science	Physics 9-12
26. Professional	Coordinator - Department of the Arts
27. Professional Educator	Principal
28. Reading	Reading
29. Secondary Education	Social Studies 9-12
30. Social Studies - Secondary/Administration	Social Studies & English
31. Spanish and French	Grades 7 & 8

TABLE 2

Recommended (Raw) Cut Scores by Item Set and Mode of Presentation

Test/Item Set		Mode: Screenprint	Mode: Computer	Diff
Reading[a]				
Set 1	Mean	13.4	13.1	.3
(Items 1-20)	SD	2.2	1.6	
	N	13	16	
Set 2	Mean	13.1	13.5	-.4
(Items 21-40)	SD	1.3	1.9	
	N	16	13	
Mathematics[b]				
Set 1	Mean	12.6	11.7	.9
(Items 1-20)	SD	2.7	2.0	
	N	16	14	
Set 2	Mean	11.7	12.4	-.7
(Items 21-40)	SD	1.5	2.1	
	N	14	16	
Writing[c]				
Set 1	Mean	14.5	13.3	1.2
(Items 1-20)	SD	1.7	3.1	
	N	8	8	
Set 2	Mean	14.3	12.6	1.7
(Items 21-40)	SD	1.8	3.0	
	N	8	8	

[a] This analysis is based on the judgments of 29 panel members: one panel member did not complete the ratings for Reading; one panel member did not follow the presentation sequence design.

[b] This analysis is based on the judgments of 30 panel members: one panel member did not follow the presentation sequence design.

[c] The statistics for the writing section are based only on the first day of data collection, as all panelists made knowledge estimation judgments based only on screenprints of writing items on the second day of data collection.

TABLE 3

Analysis of Variance Results for Mathematics Items[a]

Source of Variance	df	Mean Square	F
	Mathematics (Item Set 1)		
Day (D)	1	22.1	4.3
Order (O)	1	4.6	0.9
Mode (M)	1	6.9	1.3
D x O	1	2.2	0.4
D x M	1	13.8	2.7
O x M	1	0.5	0.1
D x O x M	1	0.0	0.0
Error	22	5.2	
	Mathematics (Item Set 2)		
Day	1	2.8	0.8
Order	1	3.0	0.8
Mode	1	3.1	0.8
D x O	1	0.0	0.0
D x M	1	3.5	0.9
O x M	1	3.1	0.8
D x O x M	1	0.0	0.0
Error	22	3.8	

[a] One panel member did not follow the presentation sequence design. Therefore, this analysis is based on judgments of 30 panel members.

TABLE 4

Analysis of Variance Results
for Reading Items[a]

Source of Variance	df	Mean Square	F
	Reading (Item Set 1)		
Day (D)	1	2.1	0.5
Order (O)	1	5.3	1.3
Mode (M)	1	0.7	0.2
D x O	1	3.3	0.8
D x M	1	3.1	0.8
O x M	1	0.3	0.1
D x O x M	1	0.4	0.1
Error	21	4.0	
	Reading (Item Set 2)		
Day	1	6.9	2.5
Order	1	2.9	1.1
Mode	1	0.7	0.3
D x O	1	0.7	0.2
D x M	1	0.1	0.0
O x M	1	0.5	0.2
D x O x M	1	0.0	0.0
Error	21	2.8	

[a] This analysis is based on the judgments of 29 panel members: one panel member did not complete the ratings for Reading; one panel member did not follow the presentation sequence design.

TABLES 5A and 5B

MATHEMATICS

THREE WAY TABLE: DAY by MODE by ORDER

RECOMMENDED CUT-SCORES for SUMS OVER ITEMS 1 - 20 MEANS, STANDARD DEVIATIONS (SD), and NUMBER of PANELISTS (N)						
	DAY 1 MODE			DAY 2 MODE		
ORDER	PRINT/ SCREEN	COMPUTER	MARGINAL	PRINT/ SCREEN	COMPUTER	MARGINAL
FIRST						
MEAN	13.6	11.2	12.4	11.0	11.3	11.2
SD	2.9	1.9	2.6	2.1	1.7	1.8
N	4	4	8	4	4	8
SECOND						
MEAN	14.6	12.6	13.6	11.2	12.0	11.4
SD	2.1	2.6	2.5	2.2	2.8	2.2
N	4	4	8	4	2	6
MARGINAL						
MEAN	14.1	11.9	13.0	11.1	11.5	11.3
SD	2.4	2.2	2.5	2.0	1.9	1.9
N	8	8	16	8	6	14

RECOMMENDED CUT-SCORES for SUMS OVER ITEMS 21 - 40 MEANS, STANDARD DEVIATIONS (SD), and NUMBER of PANELISTS (N)						
	DAY 1 MODE			DAY 2 MODE		
ORDER	PRINT/ SCREEN	COMPUTER	MARGINAL	PRINT/ SCREEN	COMPUTER	MARGINAL
FIRST						
MEAN	11.7	13.6	12.6	11.8	12.4	12.2
SD	2.3	2.1	2.3	0.5	1.0	0.9
N	4	4	8	2	4	6
SECOND						
MEAN	11.8	12.4	12.1	11.7	11.1	11.4
SD	1.3	2.8	2.1	1.5	2.2	1.8
N	4	4	8	4	4	8
MARGINAL						
MEAN	11.7	13.0	12.4	11.7	11.8	11.7
SD	1.7	2.4	2.1	1.2	1.7	1.5
N	8	8	16	6	8	14

TABLES 6A and 6B

READING

THREE WAY TABLE: DAY by MODE by ORDER

RECOMMENDED CUT-SCORES for SUMS OVER **ITEMS 1-20** MEANS, STANDARD DEVIATIONS (SD), and NUMBER of PANELISTS (N)						
	DAY 1 MODE			DAY 2 MODE		
ORDER	PRINT/ SCREEN	COMPUTER	MARGINAL	PRINT/ SCREEN	COMPUTER	MARGINAL
FIRST						
MEAN	14.6	14.1	14.3	12.9	13.2	13.0
SD	1.0	1.4	1.2	1.0	0.9	.9
N	3	4	7	4	4	8
SECOND						
MEAN	13.5	12.1	12.8	12.6	13.0	12.9
SD	3.1	1.1	2.3	4.1	2.6	2.7
N	4	4	8	2	4	6
MARGINAL						
MEAN	14.0	13.1	13.5	12.8	13.1	13.0
SD	2.4	1.6	2.0	2.0	1.8	1.8
N	7	8	15	6	8	14

RECOMMENDED CUT-SCORES for SUMS OVER **ITEMS 21-40** MEANS, STANDARD DEVIATIONS (SD), and NUMBER of PANELISTS (N)						
	DAY 1 MODE			DAY 2 MODE		
ORDER	PRINT/ SCREEN	COMPUTER	MARGINAL	PRINT/ SCREEN	COMPUTER	MARGINAL
FIRST						
MEAN	13.1	13.6	13.4	12.4	13.1	12.6
SD	0.7	2.9	2.0	1.3	2.1	1.4
N	4	4	8	4	2	6
SECOND						
MEAN	14.3	14.2	14.3	12.8	13.1	13.0
SD	1.6	2.2	1.7	1.1	0.6	0.8
N	4	3	7	4	4	8
MARGINAL						
MEAN	13.7	13.9	13.8	12.6	13.1	12.8
SD	1.3	2.4	1.9	1.1	1.0	1.1
N	8	7	15	8	6	14

TABLES 7A and 7B

MATHEMATICS

TWO WAY TABLE: DAY by MODE by ORDER

RECOMMENDED CUT-SCORES for SUMS OVER ITEMS 1-20 MEANS, STANDARD DEVIATIONS (SD), and NUMBER of PANELISTS (N)			
	DAY		
MODE	DAY 1	DAY 2	MARGINAL
PRINT SCREEN			
MEAN	14.1	11.1	12.6
SD	2.4	2.0	2.7
N	8	8	16
COMPUTER			
MEAN	11.9	11.5	11.7
SD	2.2	1.9	2.0
N	8	6	14
MARGINAL			
MEAN	13.0	11.3	12.2
SD	2.5	1.9	2.4
N	16	14	30

RECOMMENDED CUT-SCORES for SUMS OVER ITEMS 21-40 MEANS, STANDARD DEVIATIONS (SD), and NUMBER of PANELISTS (N)			
	DAY		
MODE	DAY 1	DAY 2	MARGINAL
PRINT SCREEN			
MEAN	11.7	11.7	11.7
SD	1.7	1.2	1.5
N	8	6	14
COMPUTER			
MEAN	13.0	11.8	12.4
SD	2.4	1.7	2.1
N	8	8	16
MARGINAL			
MEAN	12.4	11.7	12.1
SD	2.1	1.5	1.8
N	16	14	30

TABLES 8A and 8B

READING

TWO-WAY TABLE: DAY by MODE

RECOMMENDED CUT-SCORES for SUMS OVER ITEMS 1-20 MEANS, STANDARD DEVIATIONS (SD), and NUMBER of PANELISTS (N)			
	DAY		
MODE	DAY 1	DAY 2	MARGINAL
PRINT SCREEN			
MEAN	14.0	12.8	13.4
SD	2.4	2.0	2.2
N	7	6	13
COMPUTER			
MEAN	13.1	13.1	13.1
SD	1.6	1.8	1.6
N	8	8	16
MARGINAL			
MEAN	13.5	13.0	13.2
SD	2.0	1.8	1.9
N	15	14	29

RECOMMENDED CUT-SCORES for SUMS OVER ITEMS 21-40 MEANS, STANDARD DEVIATIONS (SD), and NUMBER of PANELISTS (N)			
	DAY		
MODE	DAY 1	DAY 2	MARGINAL
PRINT SCREEN			
MEAN	13.7	12.6	13.1
SD	1.3	1.1	1.3
N	8	8	16
COMPUTER			
MEAN	13.9	13.1	13.5
SD	2.4	1.0	1.9
N	7	6	13
MARGINAL			
MEAN	13.8	12.8	13.3
SD	1.9	1.1	1.6
N	15	14	29